Journal of Biomedical Informatics 44 (2011) 42–47

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

A set of ontologies to drive tools for the control of vector-borne diseases

Pantelis Topalis [a], Emmanuel Dialynas [a], Elvira Mitraka [b], Elena Deligianni [a], Inga Siden-Kiamos [a], Christos Louis [a,b,*]

[a] Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, 711 10 Heraklion, Crete, Greece
[b] Department of Biology, University of Crete, 711 10 Heraklion, Crete, Greece

ARTICLE INFO

Article history:
Received 20 October 2009
Available online 2 April 2010

Keywords:
Anatomy
Database
Decision support system
Insecticide resistance
Malaria
Mosquito
Tick
Transmission
Arthropod vector

ABSTRACT

We are developing a set of ontologies dealing with vector-borne diseases as well as the arthropod vectors that transmit them. After building ontologies for mosquito and tick anatomy we continued this project with an ontology of insecticide resistance followed by a series of ontologies that describe malaria as well as physiological processes of mosquitoes that are relevant to, and involved in, disease transmission. These will later be expanded to encompass other vector-borne diseases as well as non-mosquito vectors. The aim of the whole undertaking, which is worked out in the frame of the international IDO (Infectious Disease Ontology) project, is to provide the community with a set of ontological tools that can be used both in the development of specific databases and, most importantly, in the construction of decision support systems (DSS) to control these diseases.

1. The problem of vector-borne diseases

Epidemiologists have brought together in one "functional" group a series of diseases of different aetiology and pathogenesis that share one key component: their mode of transmission (see [1] and several chapters of [2] for specific questions addressing insect-borne diseases and their vectors). These diseases are transmitted by the bite of a specific arthropod vector, which is usually (but not exclusively) an insect. The pathogenic agent is usually passed with the saliva transferred during the bite to the potential patient. Two additional characteristics are shared by most vector-borne diseases, namely most people affected live in the tropical regions of the world and, connected to this, the diseases affect mostly populations that are also heavily affected by poverty. The pathogens responsible for these diseases are very diverse, ranging from protozoan parasites (e.g. *Plasmodium* spp. in malaria, *Leishmania* spp. in leishmaniosis) and bacteria (e.g. *Borrelia* spp. in Lyme disease), to worms (e.g. Nematodes in lymphatic filariasis and river blindness) and viruses (e.g. Dengue, Yellow fever). Similarly, the vectors that transmit them are also very diverse and range from mosquitoes (e.g. malaria and Dengue) and flies (e.g. Tsetse in African trypanosomiasis) to kissing bugs (Chagas' disease) and ticks (e.g. Lyme disease). This diversity is shown in Table 1, which lists several diseases along with the corresponding aetiologic agents and arthropod vectors. The great variation in the biology of both pathogens and vectors, and the ensuing differences in the illnesses caused, makes it impossible to address vector-borne diseases as a cohesive clinical entity. Importantly, these difficulties also affect significant aspects such as prevention, epidemiology, therapy, etc.

A common theme, which in a sense unites these diseases, is the fact that their transmission can be blocked if the agents that transmit them, i.e. the arthropod vectors, are removed from the pertinent chain of events [3]. Vector control has therefore historically become a *conditio sine qua non* for the control of these infections [4,5], and this fact has been exemplified by the elimination of malaria from most non-tropical areas of the globe [6]. While leading to about half a billion cases in the tropics and still being responsible for anything between one and three million deaths (mostly children in sub-Saharan Africa) every year, this killer illness has practically disappeared from Europe and North America through intense insecticidal measures aimed at eliminating the anopheline vectors [6]. It should be stressed here that, with the exception of the Yellow fever [7], no vaccine is currently available for any vector-borne disease as an alternative prevention strategy that would act on a different level than that of the actual vector. Prevention focused on the vector includes not only control of insect populations through environmental management or the use of chemicals, but also the protection of individuals through the use of clothing, repellents, nets and screens [8]. In addition, prevention

* Corresponding author at: Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, N. Plastira 100, 700 13 Heraklion, Greece. Fax: +30 281 0391104.
E-mail address: louis@imbb.forth.gr (C. Louis).

1532-0464/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2010.03.012

Table 1
Some important vector-borne diseases, their pathogens and their arthropod vectors; commonly used synonyms are also listed.

Disease	Pathogen	Vector(s)
(i) Bacterial diseases		
Louse-borne relapsing fever	*Borrelia recurrentis*	Louses
Lyme disease	*Borrelia burgdorferi*	Ticks
Plague	*Yersinia pestis*	Fleas
Tick-borne relapsing fever	var. spirochetes	Ticks
Tularemia	*Francisella tularensis*	Ticks, deer flies
(ii) Viral diseases		
Chikungunya fever	Chikungunya virus	*Aedes aegypti*, *Ae. albopictus*
Dengue fever	DENV	*Ae. aegypti*, *Ae. albopictus*
Eastern equine encephalitis	EEEV	*Aedes* spp., *Coquillettidia* spp. *Culex* spp.
Japanese encephalitis	JEV	*Culex tritaeniorhynchus*
La Crosse encephalitis	La Crosse virus	*Ae. triseriatus*
Saint Louis encephalitis	SLE	*Culex* spp.
West Nile encephalitis	WNV	*Culex* spp.
Western Equine Encephalitis	WEEV	Various mosquito species
Yellow fever	YFV	*Aedes* spp., *Haemagogous* spp.
(iii) Parasitic diseases		
African trypanosomiasis[1]	*Trypanosoma brucei*	*Glossina* spp.[2]
American trypanosomiasis[3]	*Trypanosoma cruzi*	*triatominae*[4]
Leishmaniasis	*Leishmania* spp.	*Lutzomyia* spp.[5], *Phlebotomus* spp.[5]
Lymphatic filariasis[6]	*Wuchereria bancrofti*, *Brugia malayi*	Various mosquito species
Malaria	*Plasmodium* spp.	*Anopheles* spp.
Onchocerciasis[7]	*Onchocerca volvulus*	*Simullium* spp.[8]

Commonly used synonyms: [1]Asian tiger mosquito, [2]sleeping sickness, [3]Tsetse, [4]Chagas' disease, [5]kissing bugs, [6]sand flies, [7]elephantiasis, [8]river blindness, and [9]black flies.

is complemented, in cases in which this is possible, by the use of drugs that block infection in its very initial stages [9].

Although greatly successful in the previous century, insect-control programmes are now immensely obstructed by a variety of factors. These range from community opposition to the widespread use of chemicals [10], to the development of resistance against these very chemicals by the insect vectors to be controlled [11]. Moreover, these problems are aggravated by several additional facts: resistance against drugs is also encountered in the pathogens [12]; vaccine development, if at all possible, is slow [13]; new drug development is not only slow but extremely expensive and the areas affected by the diseases in question are certainly not the ones that can easily spearhead such efforts due to the lack of economic and scientific resources in them [14]. All of the difficulties addressed above have led to a resurgence of vector-borne diseases, which now pose again a threat to more than just the tropical regions of the world [15]. It is therefore of utmost importance to develop innovative strategies for the control of vector-borne diseases. One novel approach is to use information technologies (IT) as a complement to the application of modern biochemical/biological techniques, often based on molecular biology, in the study of the biology of disease vectors. While the latter approaches make use of scientific research products such as whole genome sequences [16,17], transgenesis [18], and the use of other "intelligent" approaches [19], the former can introduce new specific tools, such as databases and DSS, that can be used for a more efficient, and often close-to-the-field management of pertinent disease data, including entomological data.

In this context, several years ago our group embarked on a long project that involves the development of ontologies dealing with vector-borne diseases and their vectors [20,21]. The obvious rationale behind this is the potential of such ontologies to unify the "language" spoken by vector biologists, epidemiologists and other specialists. It should be noted here that the usage of very specific terms and, even worse, jargon often makes it more difficult to obtain a wide understanding of certain terms. For example, the terms "refractory to" or "resistant to", combined with the words "malaria" or "*Plasmodium*" or "infection" are all synonyms. We obviously do not see the need for ontologies restricted only to the actual vectors of the vector-borne diseases but also expanding into the "area" of the diseases and, most importantly, the two have to be interoperable. The ultimate end, thus, is to build a comprehensive ontology for insect-borne diseases that may consist of sub-ontologies, each addressing a specific aspect of the whole. In the frame of the Infectious Disease Ontology project, IDO [22 and http://www.infectiousdiseaseontology.org/Home.html], we initiated this effort focusing on malaria, but we are already expanding this to encompass most other vector-borne diseases as well. The choice of developing the malaria ontology in close partnership with the IDO project was made because of its specific advantages. Having IDOMAL as an extension of such a reference ontology, as opposed to an autonomous approach, allows for a superior interoperability of all individual application ontologies in the context of the greater infectious disease domain. The ontologies that we are working on, some of which are already available and some under development, will be presented below in a summary form.

2. Ontologies and vector-borne diseases: an ephemeral account

There are several aspects of vector-borne diseases that are in need of ontological description; they range from those that deal with the diseases as such (e.g. pathogenesis, clinical aspects, therapy, etc.), to vector biology (physiological processes of the vectors) and to epidemiology and control in the widest sense of the terms (prevention, insect control, etc.). As stated earlier, these aspects are extremely diverse and complex, simply given the multitude of organisms involved (vectors and pathogens in addition to the human host) and the fact that we are often dealing with populations, rather than individuals (additional level of granularity!). The construction of a comprehensive ontology, thus, if at all feasible, must be addressed using a piecemeal approach. It is clear that certain fundamental decisions have to be taken at the initial phases, and an open-ended advance is, in our mind, a must. We therefore decided, early on, that the end product (i) would have to follow the rules set by the OBO Foundry [23] and, if no other reasons dictated a different decision, (ii) should be based on BFO, the basic formal ontology [24,25]. The reasoning behind the decision

was that if long-term interoperability of future databases and IT tools is to be achieved, these two choices are a prerequisite. We considered this choice, though, to be of more relevance to the final goal and we therefore decided to keep a certain degree of flexibility throughout the project until a unified vector-borne disease ontology is fully developed. One example for such a flexible approach is the fact that the ontology of insecticide resistance in mosquitoes, MIRO [26], which we constructed, does not comply with the BFO in its initial versions; rather, it is structured such that it can be adopted, without many problems, by the community that immediately needs to apply it in the field (see below). The MIRO forms the core of the related database on insecticide resistance (IRbase; http://anobase.vectorbase.org/ir/) that we also developed [26], and which was adopted for immediate use by the Regional Office for Africa of the World Health Organization (WHO-AFRO): all field studies that are run under direct or indirect support by WHO-AFRO are asked to submit their data to IRbase. Furthermore, it is planned to move the curation of both database and ontology to an African country with the support of WHO-AFRO. As we have not abandoned the goal of ultimately unifying all ontologies we currently construct, we are in the process of restructuring the MIRO along BFO standards, such that its contents can be later directly imported and incorporated into the comprehensive ontology on vector-borne diseases. At the same time, this will also add accuracy to the ontology (see below). A similar restructuring after the first version was made publicly available occurred with the TGMA, the ontology of the mosquito gross anatomy [20]; we decided for the reasons outlined above, that we should conform the TGMA to the CARO, the Common Anatomy Reference Ontology [27], which is BFO-based. The first version was, thus, retracted and a CARO-compliant TGMA version was then submitted to the OBO Foundry and is now available (http://www.obofoundry.org/cgi-bin/detail.cgi?id=mosquito_anatomy). In contrast to TGMA, TADS, the tick anatomy ontology that we constructed next [20, http://www.obofoundry.org/cgi-bin/detail.cgi?id=tick_anatomy], was directly built as an extension to CARO.

The MIRO has also been already submitted to, and is listed by the OBO Foundry (http://www.obofoundry.org/cgi-bin/detail.cgi?id=mosquito_insecticide_resistance). It consists of four sub-ontologies that cover all aspects of insecticide resistance of mosquito disease vectors, with a special emphasis on fieldwork and monitoring. Thus, although genetic mechanisms of resistance are covered, this is not done in detail, since many of those are processes already covered by the Gene Ontology [28,29]. Furthermore, the MIRO's fifth major component, a geographical one, uses *in toto* the controlled vocabulary Gazeteer (http://darwin.nerc-oxford.ac.uk/gc_wiki/index.php/GAZ_Project) in order to provide IRbase curators with records describing the areas in which data were collected. The MIRO is constantly being updated, upon request, by members of the international community that is involved in the study of insecticide resistance. To help cover the wishes of the pertinent community, we recruited the help of an expert on insecticide resistance who also co-authored the publication of the MIRO [26]. Moreover, all geographical locations reported to IRbase, which so far are spread over 5 continents, are annotated using the GAZ. Should a location not be listed, it is communicated to the curators of the GAZ ontology to be placed at the appropriate position. As mentioned earlier, the fact that MIRO is not BFO compliant renders it easier to be understood by non-experts, but at the same time it looses in accuracy. Table 2 illustrates this point. In MIRO, two steps are enough to define DEF (S,S,S-tributyl phosphorotrithioate), to some extent wrongly, as an insecticidal substance. In contrast, five steps are necessary in IDOMAL, but DEF is defined much more accurately here. As is the case with most, if not all biomedical ontologies, MIRO cannot be considered as complete. More insecticides are being developed, new modes of action are discovered and, unfortunately, and spread of insecticide resistance simply cannot be stopped. MIRO is therefore under steady curation and new versions are made public as soon as the community requires pertinent changes.

Table 2
The table shows the comparison between the non-BFO compliant MIRO and the corresponding term(s) after introduction into IDOMAL and adherence to BFO.

MIRO	IDOMAL
DEF *is_a* synergist	DEF *has_role* insecticide synergist
Synergist *is_a* insecticidal substance	DEF *is_a* chemical compound
	Chemical compound *is_a* abiotic object
	Abiotic object *is_a* object
	Insecticide synergist *is_a* role

Another, originally nameless, ontology that we build covers physiological processes of mosquitoes that are involved in disease transmission. Our original decision to make this an autonomous ontology was later modified, and we are presently in the process of fully incorporating it into IDOMAL (see below). The processes listed do not only address the actual disease transmission, i.e. the interplay between vectors and pathogens but, importantly, also the actual progression of events in the vector. We want to stress that the processes mentioned here are, in their vast majority, processes on the level of the organism and not cellular or sub-cellular ones, such as the ones covered by the GO [28,29]. Moreover, many of these processes are species-specific, and therefore also excluded from the GO, which is focused on processes of a general nature (but see below). Thus, (near) top level classes are, among others, behavior, sensory perception, processes of the immune system and nutrition, all physiological components that directly affect the transmission potential of disease vectors. As an example, when looking at the children terms of "behavior", one will find a line of terms leading through the adult feeding behavior, to entities such as the four phases of "interrupted feeding" (exploratory phase, imbibing phase, probing phase and withdrawal phase). The ontology also covers processes that are not directly "linked" to disease transmission and this, obviously, for reasons of completion. Because of the principle of orthogonality, as was the case with GAZ and MIRO, in all cases in which terms are already covered by established public ontologies we adhere to these, along with their descendants. For that, we search ontologies at the NCBO Bioportal (http://bioportal.bioontology.org/) and directly import relevant hits (IDs and definitions) into our ontologies. This is notably the case for the Biological Processes sub-ontology of the GO, as can be seen in Table 3 that lists a part containing metabolic

Table 3
The table shows a small part of the physiological processes of vectors, described in IDOMAL, that has extensive overlap with the GO-Biological Process sub-ontology (GO IDs are indicated in parentheses). All terms are connected with terms lying above and to the left of them with *is_a* relations. ("IDOMAL:XXXXXXX, no GO term in BP") refers to terms for which no corresponding term is found in the Biological Process sub-ontology. For three cases similar terms, indicated at the bottom part of the table, are found in the sub-ontology of molecular function.

Metabolic process (GO:0008152)
- Catabolic process (GO:0009056)
  - Carbohydrate catabolism (GO:0016052)
    - Glycolysis (GO:0006096)
    - Cleavage by carbohydrases (IDOMAL:0001299, no GO term in BP)
  - Lipid catabolic process (GO:0016042)
    - Fatty acid β-oxidation (GO:0006635)
    - Cleavage by esterases (IDOMAL:0001298, no GO term in BP)[a]
  - Protein catabolic process (GO:0030163)
    - Cleavage by peptidases (IDOMAL:0001300, no GO term in BP)[b]
      - Cleavage by serine proteases (IDOMAL:0001297, no GO term in BP)[c]
- Pigment metabolic process (GO:0042440)

GO terms describing molecular function.
[a] GO:0016788 (hydrolase activity, acting on ester bonds).
[b] GO:0008233 (peptidase activity).
[c] GO:0008236 (serine-type peptidase activity).

processes. In this example, eight of 12 terms do have corresponding terms in the GO, while for the remaining four, three have similar (but obviously not identical) terms in the molecular function sub-ontology of the GO. The decision to use terms (and their IDs) *verbatim* from ontologies that have previously found their way into the public domain is one that we strictly adhere to, as this provides one of the most crucial advantages linked to the usage of ontologies, namely the possibility of cross-talk between databases that share biological metadata. Initially we had decided against using the parent term ID, as we often did not want to import the whole tree associated to the terms. For example, in ChEBI, some insecticides are listed as acaricides, while we consider them to be bona fide insecticides. In the meanwhile we have modified this standard and decided, in most cases, to use the original IDs from OBO Foundry ontologies such as GO and GAZ. This "transcription" is now in progress and soon most such IDs will cease appearing as xrefs.

IDOMAL, is an ontology describing malaria; it is the ontology that we are in the process of populating with terms and this is the actual ontology that we decided to develop in the frame of IDO, and which we plan to expand in the near future in order to cover other vector-borne diseases as well. It is built based on BFO and the IDO reference ontology (http://www.infectiousdiseaseontology.org/IDO_files/IDO_10.08.07.obo.txt), and it is meant to cover malaria on all possible levels. More than 1800 terms currently exist in IDOMAL, even though it cannot be considered as complete. Table 4 shows semi-schematically, the contents of IDOMAL. These obviously include both the clinical aspects of the disease in the widest sense (i.e. including epidemiology, etc.) and the biology of the disease that describes processes and objects of no immediate clinical relevance. We consider as such items (e.g. proteins) involved in the penetration of both mosquito and human/vertebrate cells as well as their interacting partners in the *Plasmodium* parasites. Again, similarly to the case of the ontology of physiological processes, we have taken care to include, wherever possible, direct imports of pre-existing ontologies. This is again the case with terms already described by the GO, but an additional example is the *Plasmodium* parasite life cycle; all stages have cross-references to the, at the moment, inactive *Plasmodium* life cycle ontology.

We have now finished expanding the IDOMAL to cover the immunology of malaria. This now covers the immune responses and the immune state of the vertebrate hosts of the parasites and in particular humans, and it is planned to also include, in the future, the immune responses depicted by anopheline vectors when they are "infected" with *Plasmodium* parasites during a blood meal. While insect immunity's possible interaction with pathogens carried by the vector could be potentially used for intelligent schemes aiming at halting pathogen transmission [30], the human immune system could also be "recruited" in strategies aiming at stopping malaria [31]: it should be stressed again that no vaccines are available for malaria, and therefore any instrument that may be of help in developing them is of utmost importance.

Since both IDO and IDOMAL are still in development, even if at an advanced stage, our ontology may well have to be modified later to take care of discrepancies between the two ontologies, given the fact of their intimate relationship. Table 5 shows identical entities in the IDO and IDOMAL, which obviously share the same definition despite the fact that their names are slightly different; eventually, IDOMAL will switch to IDO's ID number, keeping the alternate name, where necessary, as a synonym. In contrast, Table 6 lists five examples of terms that, although very similar, have a clearly different meaning in the two ontologies; the IDOMAL terms, here, have a more specific meaning, thus at the end the terms will continue appearing with different ID numbers in IDO and IDOMAL.

Table 4
Semi-schematic listing of the contents of IDOMAL. The hierarchy of terms listed here does not correspond to what is to be found in IDOMAL, due to the BFO format followed. Not all classes are listed.

Biology of disease		
	Malaria immunology	
	Malaria pathogens	
	Parasite–vector interactions	
	Parasite–vertebrate interactions	
Clinical features		
	Malaria forms	
	Severe malaria	
	Cerebral malaria	
	Malaria in pregnancy	
	Malaria in children	
Diagnostic procedures		
Epidemiology		
Disease control		
	Malaria eradication	
	Vector control	
	Treatment	
		Chemoprophylaxis
		Chemotherapy
		Immunization
	Treatment of severe malaria	
Parasite biology		
	Plasmodium cycle	
	Plasmodium species	
	Drug resistance	
Vector biology		
	Anopheline species	
	Insecticide resistance	
	Mosquito immunology	
	Transmission-related physiology	
	Population biology	

3. Ontologies and vector-borne diseases: concluding remarks

The ontologies that we are constructing could be described, in a sense, as pure application ontologies that are meant to form the basis for specific tools such as specific databases or decision support systems for various diseases. The need for such tools became apparent immediately after the first working version of the MIRO and its sibling IRbase were made public. Not only did the international community, most prominently WHO-AFRO, immediately decide to adopt both tools, but also already within a few months after the initiation of data submission, there are about 1500 population records in the database. This is about 1400 more samples than what the previous insecticide resistance section in VectorBase carried, the only repository for data of this kind. In addition to databases that are driven by ontologies in an increasing fashion (see for example databases using the ontology-depending schema Chado [32], such as FlyBase [33,34] and VectorBase [35,36]), ontologies are ideal tools for the design and function of intelligent DSS. As a matter of fact, we are aware of at least two such IT tools being developed presently, the malaria decision support system MDSS (http://www.ivcc.com/projects/mdss.htm) and the Dengue decision support system (http://www.ivcc.com/projects/ddss.htm), that are driven in part, by ontologies developed or specifically adapted for that purpose. In cases such as vector-borne diseases, whose control is also hampered by weak infrastructure in endemic countries, these DSS can be used by medical workers and health agencies in remote areas, either for ongoing studies or, most crucially, in cases that need immediate attention [37,38] such as emerging epidemics.

One of the intricacies that we are already faced with is the planned expansion of the malaria-oriented ontologies, to cover many other vector-borne diseases. To understand the magnitude of the challenge one should think of the fact, as stated earlier, that

Table 5
Terms in IDO with their counterparts in the draft IDOMAL.

IDO: term, ID	IDOMAL: term, ID	Common definition
Host role, 408	Host, 0000055	A role borne by an organism by virtue of the fact it provides an environment supportive for the survival or reproduction of an entity of another type
Parasite role, 443	Parasite, 0000995	A symbiont role borne by an organism in virtue of the fact that it derives a growth, survival, or fitness advantage from symbiosis while the other symbiont's growth, survival, or fitness is reduced
Reservoir of infectious agent role, 424	Reservoir, 0000058	A role borne by a material entity by virtue of the fact that it is a habitat in which an infectious agent is persisting and multiplying and from which the infectious agent can be transmitted
Pathogen role, IDO:405	Pathogen, 0000063	A role borne by an object in virtue of the fact that it is sufficiently close to an organism towards which it has the pathogenic disposition to allow processes resulting in disorder to occur
Infectious disease, 436	Infectious disease, 000001051	A disease whose physical basis is an infection
Virulence, 466	Virulence, 0000004	A quality that inheres in an infectious agent and is the degree to which realizations of the infectious disease caused by the infectious agent become severe or fatal
Organism population, 509	Population, 0001254	An aggregate of organisms
Susceptibility, 467	Susceptibility 0001048	A quality that inheres in an entity and is the degree to which it can be harmed by another entity of a certain type
Immunization against infectious agent, 497	Immunization, 0001039	A process by which an organism acquires immunity to an infectious agent

Table 6
Terms in IDO with their counterparts in the draft IDOMAL.

IDO: term, ID	Definition	IDOMAL: term, ID	Definition
Infectious disease prevalence, 485	A quality that inheres in an organism population and is the number of realizations of an infectious disease of a certain type in the population at a specified time	Prevalence of malaria, 0000019	The number of malaria cases existing in a given population at any given time
Infectious disease incidence, 479	A quality that inheres in an organism population and is the number of realizations of an infectious disease of a certain type for which the infectious disease course begins during a specified period of time	Incidence of malaria, 0001243	Prevalence over a stated time period independent of whether the disease resulted from a new infection or not
Infectious disease epidemic, 502	A process in which there is a relatively significant increase in the infectious disease incidence of a certain type of infectious disease, relative to the endemic level of realizations of that infectious disease, in an organism population located in a geographically connected region	Epidemic malaria, 0000116	Spread of malaria across a population beyond what is characterized as endemic
Infectious disease course, 495	A disease course that is the realization of an infectious disease	Progression of malaria, 0000091	All clinical features of malaria from infection to cure or death
Herd immunity to infectious organism, 447	A collective resistance disposition that inheres in an infectious population in virtue of the fact that a sufficient number of members of the population have immunity to an infectious agent	Herd immunity, 00000352	Resistance of a group to a pathogen due to immunity of a large proportion of the group to that pathogen
Acquired immunity to infectious agent, 621	An immunity to infectious agent that inheres in an organism in virtue of lymphocytes and lymphocyte receptors that came into being as a result of a primary immune response	Acquired immunity to malaria, 0000543	Immunity to malaria gradually acquired by infection
Vaccination against infectious agent, 499	An active immunization process that begins with exposure of an organism to a vaccine and results in immunity against an infectious agent	Malaria vaccination, 0000021	The administration of antigenic material from malaria pathogen to produce immunity to malaria

vector-borne diseases represent major threats to public health in wide and ecologically diverse areas of the world, that they are caused by completely different pathogens and that completely different species of vectors transmit them. Thus, the challenge now is how to cover this broad spectrum of facts in a single ontology. There is naturally the possibility of cutting through the Gordian knot by devising separate ontologies for each disease. The counter-argument in this case would be that, brought to an extreme, each pathogen-related malaria form (i.e. tertian, malignant and benign, and quartan, which all have some distinct clinical features) should have its own ontology, similar to the different forms of lymphatic filariasis, which are caused by different species of nematodes but whose clinical aspect differ only slightly. In addition, similarities between these diseases and the agents that transmit them may be obscured if different ontologies were used, and this would certainly have a negative impact on their value in the long term. Therefore, we are still trying to solve the knot in a non-Alexandrian way. By attempting to merge the ontologies *in spe* into one, we can also actively support the rules of the OBO Foundry and provide an example of how the construction of a large and comprehensive ontology can, later on, provide advantages to its users.

Acknowledgments

The work was funded by contract HHSN266200400039C from the National Institute of Allergy and Infectious Diseases in the frame of the VectorBase project and by the BioMalPar and EViMalR European Networks of Excellence supported by European Grants (LSHP-CT-2004-503578 and HEALTH-F3-2009-242095) in the frame of the 6th and 7th Framework Programmes. The authors would like to thank numerous colleagues who helped at different stages of the work, and in particular Drs. John Vontas for his contribution to the MIRO, Frank Collins for his encouragement and support in the frame of VectorBase and Barry Smith and Lindsay Cowell for hosting us in the IDO community.

References

[1] Goddard J. Infectious diseases and arthropods. Totowa, NJ: Humana Press; 2000.
[2] Marquardt WC, Kondratieff BC. Biology of disease vectors. 2nd ed. Burlington, MA: Elsevier Academic Press; 2005.

[3] Hemingway J, Beaty BJ, et al. The innovative vector control consortium: improved control of mosquito-borne diseases. Trends Parasitol 2006;22:308–12.
[4] della Torre A, Arca B, et al. The role of research in molecular entomology in the fight against malaria vectors. Parassitologia 2008;50:137–40.
[5] Peter RJ, Van den Bossche P, et al. Tick, fly, and mosquito control – lessons from the past, solutions for the future. Vet Parasitol 2005;132:205–15.
[6] de Zulueta J. The end of malaria in Europe: an eradication of the disease by control measures. Parassitologia 1998;40:245–6.
[7] Roukens AH, Visser LG. Yellow fever vaccine: past, present and future. Expert Opin Biol Ther 2008;8:1787–95.
[8] Hill J, Lines J, Rowland M. Insecticide-treated nets. Adv Parasitol 2006;61:77–128.
[9] Greenwood BM. Control to elimination: implications for malaria research. Trends Parasitol 2008;24:449–54.
[10] Schapira A. DDT: a polluted debate in malaria control. Lancet 2006;368:2111–3.
[11] Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 2000;45:371–91.
[12] Laufer MK. Monitoring antimalarial drug efficacy: current challenges. Curr Infect Dis Rep 2009;11:59–65.
[13] Langhorne J, Ndungu FM, et al. Immunity to malaria: more questions than answers. Nat Immunol 2008;9:725–32.
[14] Craft JC. Challenges facing drug development for malaria. Curr Opin Microbiol 2008;11:428–33.
[15] Gubler DJ. Resurgen vector-borne diseases as a global health problem. Emerg Infect Dis 1998;4:442–50.
[16] Holt RA, Subramanian GM, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 2002;298:129–49.
[17] Nene V, Wortman JR, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 2007;316:1718–23.
[18] James AA. Preventing the spread of malaria and dengue fever using genetically modified mosquitoes. J Vis Exp 2007:231.
[19] Rasgon JL. Using predictive models to optimize Wolbachia-based strategies for vector-borne disease control. Adv Exp Med Biol 2008; 627:114–25.
[20] Topalis P, Tzavlaki C, et al. Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. Insect Mol Biol 2008;17:87–9.
[21] Topalis P, Lawson D, Collins FH, Louis C. How can ontologies help vector biology? Trends Parasitol 2008;24:249–52.
[22] Cowell LG, Smith B. Infectious disease ontology. In: Sintchenko V, editor. Infectious disease informatics. New York: Springer; 2010. p. 373–96.
[23] Smith B, Ashburner M, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;25:251–5.
[24] Simon J, Dos Santos M, Fielding J, Smith B. Formal ontology for natural language processing and the integration of biomedical databases. Int J Med Inform 2006;75:224–31.
[25] Grenon P, Smith B, Goldberg L. Biodynamic ontology: applying BFO in the biomedical domain. Stud Health Technol Inform 2004;102:20–38.
[26] Dialynas E, Topalis P, et al. MIRO and IRbase: IT tools for the epidemiological monitoring of insecticide resistance in mosquito disease vectors. PLOS Negl Trop Dis 2009;3(6):e465. doi:10.1371/journal.pntd.000046.
[27] Haendel MA, Neuhaus F, Osumi-Sutherland D, Mabee P, Mejino Jr JLV, Mugall CJ, et al. CARO – the common anatomy reference ontology. In: Burger A, Davidson D, Baldock R, editors. Anatomy ontologies for bioinformatics: principles and practice. New York: Springer; 2008. p. 327–50.
[28] Ashburner M, Lewis S. On ontologies for biologists: the Gene Ontology – untangling the web. Novartis Found Symp 2002;247:66–80. discussion 80–3, 84–90, 244–52.
[29] Harris MA, Clark J, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258–61.
[30] Christophides GK, Vlachou D, Kafatos FC. Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*. Immunol Rev 2004;198:127–48.
[31] Pierce SK, Miller LH. World Malaria Day 2009: what malaria knows about the immune system that immunologists still do not. J Immunol 2009;182(9):5171–7.
[32] Mungall CJ, Emmert DB. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics 2007;23:i337–46.
[33] Gelbart WM, Crosby M, et al. FlyBase: a *Drosophila* database. The FlyBase consortium. Nucleic Acids Res 1997;25:63–6.
[34] Tweedie S, Ashburner M, et al. FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Res 2009;37:D555–9.
[35] Megy K, Hammond M, et al. Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. Infect Genet Evol 2009;9(3):308–13.
[36] Lawson D, Arensburger P, et al. VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res 2009;37:D583–7.
[37] Thomson MC, Connor SJ, et al. The ecology of malaria – as seen from Earth-observation satellites. Ann Trop Med Parasitol 1996;90:243–64.
[38] Coleman M, Sharp B, et al. Developing an evidence-based decision support system for rational insecticide choice in the control of African malaria vectors. J Med Entomol 2006;43:663–8.